The Efficiency of Sequence-Specific Separation of DNA Mixtures for Biological Computing

Julia Khodor and David K. Gifford
Laboratory for Computer Science, Massachusetts Institute of Technology

Abstract

We report a series of experimental observations on the efficiency and fidelity of sequence-specific DNA extraction operations. We examined the solution-based annealing of bead-bound probes to target molecules followed by magnetic separation of the annealed probe-target complexes from solution. Our experiments measured how efficiently a 20-mer probe could be used to purify a 40-mer radiolabeled single-stranded target with varying target concentrations. Our results suggest that with perfectly homologous probes and targets, recovery rates of 8-24% can be expected, and when a target is not homologous to the probe less than 1% of the target is recovered. Because our conditions were designed to be ideal, the rates we obtained may represent an upper bound on the efficiency of sequence-specific separation of DNA mixtures using magnetic beads.

1. Introduction

A central primitive operation for DNA computing is the ability to extract molecules that contain a desired DNA sequence from solution. One application of extract is to apply constraints to a mixture of molecules that represent possible answers to a problem to find one or more answer molecules that meet the desired constraints.

Two well known sequence based affinity purification techniques are affinity column separation, where probe DNA is immobilized on solid support, and magnetic bead separation, where probe DNA is free to circulate in solution until the magnetic field is applied, drawing the magnetic beads and probe-target complexes attached to the beads to one side of the test tube.

Most laboratory work in DNA computing [1, 4] has used magnetic bead separation technology. In this previous work the product of each separation step was amplified by PCR before the application of the next separation step. Later theoretical work, such as the sticker model and the refinery model [5], proposed the use of separation without the subsequent use of PCR. The feasibility of using separation techniques without intervening amplification depends on whether the separation techniques satisfy certain assumptions, such as 1) high efficiency of separation (low false negative rate); 2) high fidelity of separation (low false positive rate); 3) the efficiency and fidelity of separation do not significantly degrade with the change in concentration of the target molecule; and 4) the loss of materials through the separations is negligible. The last two assumptions are particularly important in the context of the refinery model. While no claims on the separation efficiency have been made by the manufacturers of the magnetic

beads, the DNA computing literature has assumed 90 to 95% efficiency. To the best of our knowledge, there exists no published quantitative study of the efficiency of the magnetic bead separation.

In this paper we investigate the efficiency of magnetic bead separation under a number of experimental conditions. We discuss our choice of reaction conditions and provide a short summary of our experimental protocol (Section 2), present the results of the experiments (Section 3), discuss the implications of our experimental results to computing paradigms (Section 4), and conclude that amplification steps may be needed in certain DNA computation systems if separation technology is not improved beyond our experimental observations (Section 5).

2. Materials and Methods

In the experiments described below, we used Dynabeads M-280 streptavidin beads and a Magnetic Particle Concentrator (MPC) from Dynal, Inc. All oligonucleotides were manufactured by Gibco BRL.

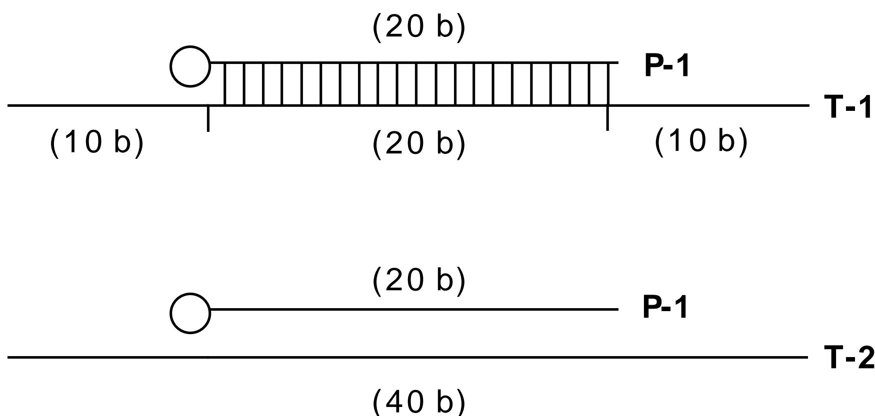

Figure 1. *Match patterns between the bead-attached probe P-1 and free-floating targets T-1 and T-2.*

Our goal in designing these experiments was to create an ideal set of conditions, so as to discover the upper bounds on the efficiency of magnetic bead separation. To that end we designed a 20-mer probe P-1, biotinylated on the 5′ end. To minimize effects of steric hindrance, we designed our target molecules to be only 40 bases long. As shown in Figure 1, T-1 is a 40-mer, whose middle 20 bases are an exact match for P-1, while T-2 is a 40-mer with no appreciable matches with P-1. We also manufactured compliments of T-1 and T-2, called respectively T-1C and T-2C. To make sure secondary structure played no role in

the behavior of the system, all five oligonucleotides were designed to have no appreciable secondary structure, and were tested for inappropriate intramolecular interactions.

We chose reaction conditions that are similar to those that would be encountered by Kaplan [4]. We used bead-bound P-1 at 10^{-8} molar, and target molecules at 10^{-10} molar and 10^{-12} molar.

The experiments consisted of five steps that were performed according to the instructions supplied by Dynal, Inc. We first mix solutions of ^{32}P-labeled (*) T-1 or T-2 with P-1-coated magnetic beads in 6X SSC buffer. We allow 30 minutes for the reactions to hybridize. Second, we apply a magnetic field, and remove the supernatant into a "-" (negative) scintillation counter (SC) vial. Third, we wash the pellet three times in 6X SSC, each time removing the supernatant into the "-" SC vial. Fourth, after adding water and heating at 67C for 5 minutes, we apply a magnetic field again, this time removing the supernatant into a "+" (positive) SC vial. The fifth and final step is to resuspend the pellet in water again and to move the resuspension into an "L" (loss) SC vial. All vials are then counted in the SC counter for 9 minutes, or until a reasonable level of certainty is achieved.

In the case where we use *T-1, we can measure the rate of false negatives by measuring the amount of *T-1 eluted into the "-" vial. In case where we use *T-2, we can measure the rate of false positives by measuring the amount of *T-2 eluted into the "+" vial. In both cases we can measure the rate of loss by measuring the amount of radiolabeled target which ended up in the "L" vial. Since the probe is in molar excess of at least 100x over the target in all experiments, we expect our single-target experiments to provide error rates reasonably close to multiple-target experiments. Since we know single-stranded DNA to be rather non-specific, in that it prefers a bad match to no match, we also performed a series of experiments where a labeled target was supplemented with an equal concentration of its cold complement.

3. Results

We report on 13 separation experiments. Experiments designed to determine false negative rates were performed, as discussed in Section 2, for two different concentrations with *T-1 by itself (repeated twice for the higher concentration and three times for the lower), and with only the higher concentration in the presence of T-1C molecules (repeated three times). Those experiments aimed at determining the false positive rates were performed only with the higher concentration of the target, both with (repeated three times) and without (repeated twice) T-2C present. The value of the baseline SC counts was obtained by averaging the values of 46 runs of blank samples (samples that only contain SC fluid).

Target	Tube	Concentration (M)	Positive	Negative	Loss	Control	Baseline
*T-1	1	1.00E-10	118267.28	22393.66	4626.97	944111.06	616.61
	3		173159.12	22236.70	5253.00		
	4	1.00E-12	1668.57	1088.29	608.15	9076.52	
	5		1797.89	918.35	634.90		
	6		2155.94	931.33	632.50		
*T-2	7	1.00E-10	1231.79	43897.46	1225.59	1010867.81	
	8		4702.60	44970.43	960.18		
	9		1611.17	43466.70	822.66		
*T-1 & T-1C	13	1.00E-10	107101.15	19034.06	6652.16	944111.06	
	14		96559.67	11390.36	4836.40		
	15		175445.78	20507.44	12232.26		
*T-2 & T-2C	19	1.00E-10	5862.52	38908.08	2242.21	1010867.81	
	20		3327.71	38727.52	668.99		

Table 1. *Raw data from the separation experiments. Values in columns "Positive," "Negative," and "Loss" are the SC counter readings of the corresponding vials (see Section 2). Values in the "Control" column are the SC counter readings of the same amount of radioactive target as was put in the corresponding reaction. Notice that values in the "Negative" column represent only 1/25th of the content of the "-" vial (see Section 2).*

Target	Tube	Concentration (M)	Positive	Negative	Loss	Control	Total Counts Pos+Neg+Loss	Accounted for
*T-1	1	1.00E-10	117650.67	544426.25	4010.36	943494.45	666087.28	70.60%
	3		172542.51	540502.25	4636.39		717681.15	76.07%
	4	1.00E-12	1051.96	11792.00	0.00	8459.91	12843.96	151.82%
	5		1181.28	7543.50	18.29		8743.07	103.35%
	6		1539.33	7868.00	15.89		9423.22	111.39%
*T-2	7	1.00E-10	615.18	1082021.25	608.98	1010251.20	1083245.41	107.23%
	8		4085.99	1108845.50	343.57		1113275.06	110.20%
	9		994.56	1071252.25	206.05		1072452.86	106.16%
*T-1 & T-1C	13	1.00E-10	106484.54	460436.25	6035.55	943494.45	572956.34	60.73%
	14		95943.06	269343.75	4219.79		369506.60	39.16%
	15		174829.17	497270.75	11615.65		683715.57	72.47%
*T-2 & T-2C	19	1.00E-10	5245.91	957286.75	1625.60	1010251.20	964158.26	95.44%
	20		2711.10	952772.75	52.38		955536.23	94.58%

Table 2. *Adjusted data from the separation experiments. Data from* Table 1 *has been adjusted for background and scaled ("Negative" column). Total counts (the sum of adjusted and scaled "Positive," "Negative," and "Loss" values), and the percentage of control values they represent have also been calculated.*

Table 1 shows the raw SC counts from the experiments, while Table 2 shows data adjusted for the base counts. Notice that the Loss counts for tube 4 in Table 1 are below the background. Since radioactive decay is a Poisson process, its error function is the square root of its value [3]. The amount by which Loss counts in tube 4 are below the background is within the square root error and,

indeed, within the 10% error commonly considered acceptable. We therefore replace the negative number that would have appeared in the tube 4 Loss space in Table 2 with 0.

Target	*Tube*	*Concentration* (M)	*Recovery, p* Pos/Total	*Negative, q* Neg/Total	*Negative error* Neg/Total	*Positive error* Pos/Total	*Loss* Loss/Total	*False negative, e-* N-error+Loss	*False positive, e+* P-error+Loss
*T-1	1	1.00E-10	17.66%		81.73%		0.60%	82.34%	
	3		24.04%		75.31%		0.65%	75.96%	
	4	1.00E-12	8.19%		91.81%		0.00%	91.81%	
	5		13.51%		86.28%		0.21%	86.49%	
	6		16.34%		83.50%		0.17%	83.66%	
*T-2	7	1.00E-10		99.89%		0.06%	0.06%		0.11%
	8			99.60%		0.37%	0.03%		0.40%
	9			99.89%		0.09%	0.02%		0.11%
*T-1 & T-1C	13	1.00E-10	18.59%		80.36%		1.05%	81.41%	
	14		25.97%		72.89%		1.14%	74.03%	
	15		25.57%		72.73%		1.70%	74.43%	
*T-2 & T-2C	19	1.00E-10		99.29%		0.54%	0.17%		0.71%
	20			99.71%		0.28%	0.01%		0.29%

Table 3. *Efficiency, fidelity, and error rates of magnetic bead separation have been calculated from the experimental data in* Table 2.

Finally, Table 3 shows the calculated efficiencies and error rates for the above experiments. It is convenient to define several terms at this point. In a separation reaction, let p be the probability that a match is identified as such; q the probability that a mismatch is identified as such; $e-$ false negative rate; $e+$ false positive rate.

It is easy to observe from the experimental data that p, q, $e-$, and $e+$ all change with the variations in the target concentration. Adding complement of the target to the reaction seems to have a small effect on the values of interest. More experiments are needed to precisely characterize the effect of adding the complement, or any other DNA strand, on the behavior of a magnetic bead separation reaction.

4. Discussion

Our experimental data indicates that the efficiency and fidelity of the magnetic bead separation reactions are dependent on the initial concentration of target molecules. Furthermore, the experimental data indicates that while fidelity (q) of the protocol is over 99%, the efficiency (p) is between 8 and 24%.

Low separation efficiency, p, may be troublesome in the context of models such as the sticker model [5], which rely on performing a large number of sequential separations without intervening DNA amplification steps. While the

fidelity of separation, q, is high, sequential separations to increase fidelity will result in substantial target losses.

For example, if it is desired to have more probe-homologous "good" target molecules after the separation than non-homologous "bad" target molecules, then multiple separation steps may be required. Multiple separation steps will be required when the starting ratio of "good" to "bad" targets is less then $(1-q)/p$. Suppose that original mixture consists of M molecules altogether: G good, and B bad. Then after the separation, the tube labeled "good" will contain $M' = pG + (1-q)B$ molecules, of which pG will be good, and $(1-q)B$ will be bad. As we have assumed above that $G/B < (1-q)/p$, then M' will consist of $<(1-q)B$ good molecules, and $(1-q)B$ bad molecules. This would necessitate the use of sequential separations which would further decrease the yield of the protocol, necessitating the use of DNA amplification.

In Section 1 we outlined some of the basic assumptions made by the models which propose to employ separations with no intervening DNA amplification steps. Our results suggest that 1) p has a moderate value, which may require DNA amplification between separation steps; 2) q is relatively high, but can still present problems due to moderate p; 3) p and q appear concentration-dependent; 4) the loss of the target molecules is appreciable. The last point is important, since it restricts the number of cascaded separations, and, in the case of the refinery model [5], the width of the compound separator unit.

The refinery model will need to be analyzed further if it is based on the magnetic bead separation [3, 5]. This is due to the apparent dependency of efficiency, p, and fidelity, q, of separation on the concentration of the target molecules. This dependency, combined with the moderate value of p (which would force the drops in the concentration of target), would necessitate the use of the generating functions in the probability calculations of the refinery model. Also, the fact that p and q are vastly different would force the probability calculation formulas in the refinery model to be expanded to reflect that difference.

As discussed in Amos, et al. [2], some authors that analyze the complexity of DNA computing do not take into account all of the steps that are required to perform a computation, and thus their analyses underestimate expected completion times. Based on our experimental data, complexity measures of DNA computations may need to account for additional steps of DNA amplification if sequence of separations is employed.

5. Conclusions

We presented the experimental data that show that in magnetic bead separation one can expect efficiency of 8 to 24%, and fidelity of more then 99%. We have observed that the efficiency and fidelity of the separation depend on target

concentrations. Based on our results, it appears that DNA amplification, such as PCR, may be required in certain types of DNA computing applications.

Acknowledgments

This paper has its origins in the class project for *Computer Science and Molecular Biology*, taught at MIT by Prof. David K. Gifford and Prof. Bonnie Berger. We would like to thank the rest of the team which worked on the class project—Alexander Hartemink, Brad Lichtenstein, Mariya Minkova, and Peter Szilagyi, all of Massachusetts Institute of Technology. The authors would also like to thank Dr. Brenton R. Graveley of Harvard University for his expert advice on magnetic bead separation, and Prof. Douglas Melton of Harvard University for providing laboratory space and resources for performing experiments described in this paper. In addition, Khodor gratefully acknowledges the support of the Genomics Training Grant.

References

[1] Leonard Adelman. Molecular computation of solutions to combinatorial problems. *Science*, 266:1021-1024, 1994

[2] Martyn Amos, Alan Gibbons, and Paul E. Dunne. The complexity and viability of DNA computations (Extended draft)

[3] William Feller. *An Introduction to Probability Theory and Its Applications, Volume I, 3rd edition*. John Wiley & Sons, 1968.

[4] Peter Kaplan, Guillermo Cecchi, Albert Libchaber. Molecular computation: Adelman's experiment repeated, 1995.

[5] Sam Roweis, Erik Winfree, Richard Burgoyne, Nickolas V. Chelyapov, Myron F. Goodman, Paul W. K. Rothemund, Leonard M. Adelaman. A sticker based model for DNA computation, 1996.